Josai Mathematical Monographs
vol. 8 (2015), pp. 73-83

Eulerian distribution with a missing number

Takahiro TSUCHIYA

Abstract. We consider a discrete distribution induced by the sorting algorithm of modified bucket sort. The systematic numbers appear in this sorting process. The discrete probability distribution for the numbers is called Eulerian distribution. In this paper, the recurrence relation for Eulerian distribution with a missing number is derived. The moment of the distribution is also given by the recurrence relation.

1. Introduction

Bucket sort, also called bin sort, is a sorting algorithm that performs by dividing an array into a number of buckets. The computational time of average case performance is $O(n+k)$, where n is a size of data and k is the number of different buckets. The sorting method is relatively faster than other sorting algorithms. Tsuchiya and Nakamura (2007, 2009) investigated a distribution of the number of buckets in the modified bucket sorting process that does not determine the number of buckets or bins. The distribution is called "Eulerian distribution". The number of buckets depends on the initial state of the data or the sequence of the numbers. Tsuchiya and Nakamura (2009) also provided the sequence of numbers represented by the recurrence relation so called "Eulerian numbers" (Euler, 1755 ; Graham et al., 1994). The numbers appear in the derivation process of the discrete distribution.

The aim of this paper is to derive the recurrence relation for Eulerian distribution with a missing number.

In the next section, the algorithm of the modified bucket sort is described. The formula of the recurrence relation for the discrete distribution is shown in Section 3. The recurrence relation for the moment of the distribution is given in Section 4. We consider Eulerian distribution with a missing number in Section 5. Numerical comparisons are also given in Section 6. Finally, we conclude the paper with a brief discussion in Section 7.

2. Algorithm of modified bucket sorting

We have a deck of well shuffled n cards numbered from 1 to n. The following procedure is carried out in order to sort a deck of cards in ascending order.

Sorting Procedure

STEP 1 If the top of a deck of cards on hand is a card number k and a card number $k+1$ is on the table, then the card k is put on the card $k+1$.

STEP 2 If the card $k+1$ is not on the table, then the card k is not put on any cards on the table but is put on the surface of the table.

STEP 3 STEP 1 and 2 are repeated until the cards on hand disappear.

STEP 4 There are m set of cards on the table ($1 \leq m \leq n$).

STEP 5 The sorting is completed by bundling up in ascending order of the top of the cards.

The differences between ordinary bucket sort and the above sort are as follows. In the modified bucket sort, (1) it is assumed that there is no same card number from 1 to n in n cards, (2) it is not known how many buckets we need, and (3) if the card $k+1$ is on the table, then the card k is put on it. The maximum number of "bunches or stacks of cards" finally obtained in STEP 4 is n, and the minimum is 1. Our interest is the number of bunches of cards in STEP 4. We note here that the number of bunches of cards corresponds to that of buckets or bins in bucket sorting.

For example, when $n = 4$, if the arrangement of cards is $(3, 1, 2, 4)$, then the number of bunches of cards is 3. Tsuchiya and Nakamura (2009) have given the recurrence relation for the probability distribution of the number of bunches of cards, called Eulerian distribution.

In this paper, we derive the recurrence relation for Eulerian distribution with a missing number. The recurrence relation for the moment of the distribution is also given.

3. Distribution of the number of bunches of cards (Eulerian distribution)

Let X be the random variable which denotes the number of bunches of cards and let $M_n(i)$ be the total number of case that $X = i$ in n cards. Then the following recurrence relation holds (Tsuchiya and Nakamura, 2009):

$$\begin{cases} M_n(1) = M_n(n) = 1 & (n \geq 1) \\ M_n(i) = iM_{n-1}(i) + \{n-(i-1)\}M_{n-1}(i-1) & (n \geq 3,\ i = 2, \cdots, n). \end{cases} \tag{1}$$

The values of $M_n(i)$ ($1 \leq i \leq n \leq 10$) are given by Table 1. Hereafter "bunches of cards" is simply called "bunches".

Table 1. $M_n(i)$ $(n = 1, 2, ..., 10)$

$n\backslash i$	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	1	4	1							
4	1	11	11	1						
5	1	26	66	26	1					
6	1	57	302	302	57	1				
7	1	120	1191	2416	1191	120	1			
8	1	247	4293	15619	15619	4293	247	1		
9	1	502	14608	88234	156190	88234	14608	502	1	
10	1	1013	47840	455192	1310354	1310354	455192	47840	1013	1

The formula is equivalent to that of Eulerian numbers by Euler (1755). $M_n(i)$ and Eulerian numbers $A(n, k)$ are exactly related with $M_n(i) = A(n, i-1)$ for $k = i - 1$ (Tsuchiya and Nakamura, 2009). Then, the probability that $X = i$ in n cards can be obtained as $P_n(i) = M_n(i)/n!$ and is given by the following recurrence formula:

$$\begin{cases} P_n(1) = P_n(n) = \dfrac{1}{n!} & (n \geq 1) \\ P_n(i) = \dfrac{i}{n}P_{n-1}(i) + \dfrac{n-(i-1)}{n}P_{n-1}(i-1) & (n \geq 3,\ i = 2, \cdots, n). \end{cases} \tag{2}$$

The probability distribution represented by the recurrence formula (2) is called Eulerian distribution. Figure 1 shows the probability distribution $P_n(i)$ for $n = 5, 10, 20, 30, 50$ and 100. We see that the distribution is symmetric with respect to the mean. It is known that the convergence to normal distribution is very fast (Tsuchiya and Nakamura, 2009; Nakamura, 2015).

4. Moment of Eulerian distribution and recurrence relation

In this section, we introduce the recurrence relation of the moment, the expectation and the variance of Eulerian distribution. The recurrence relation of the moment about the mean was given by Mann (1945), but here that of the moment about the origin is presented. A general term of the expectation and the variance are induced using the recurrence relation.

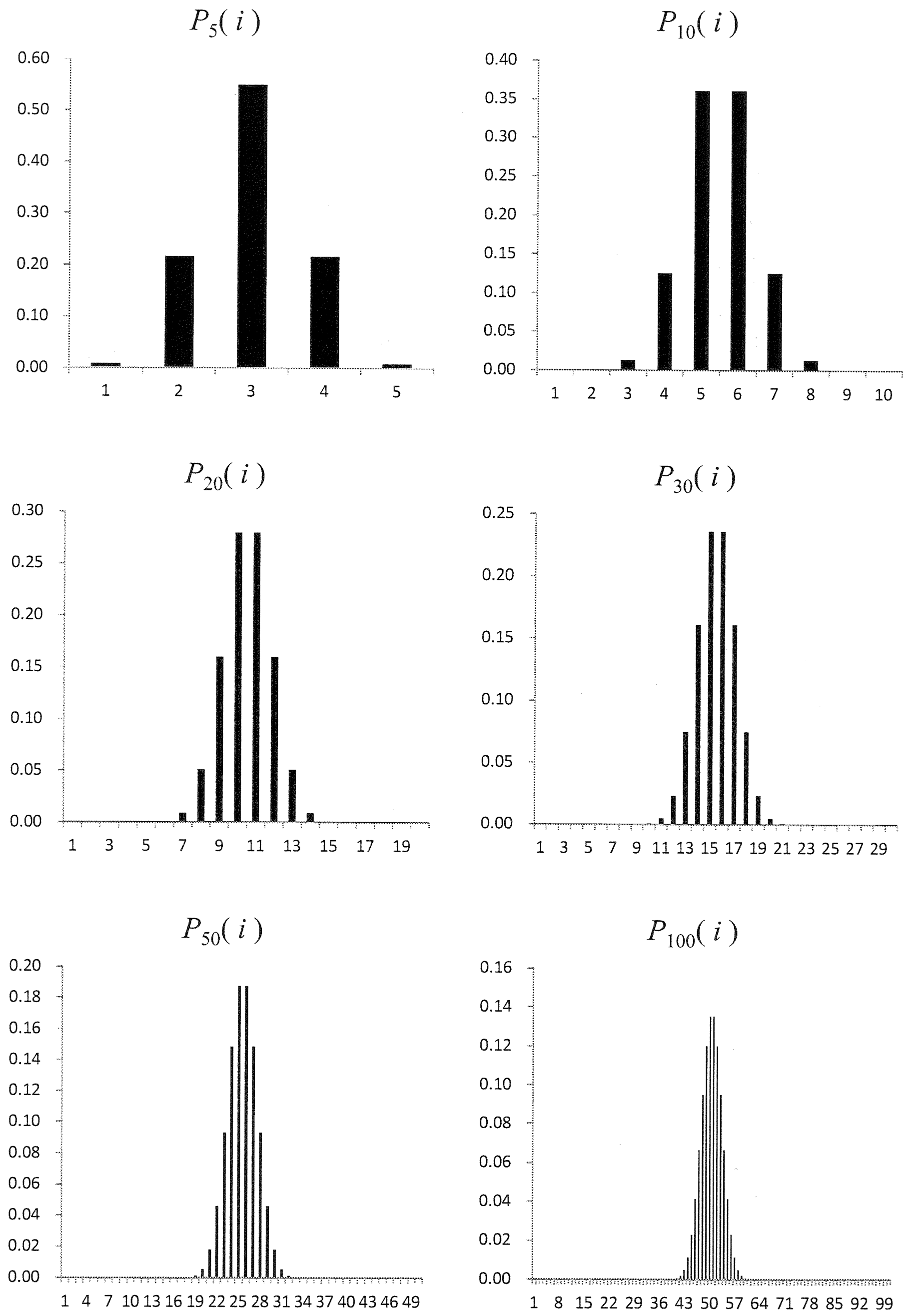

Figure 1. $P_n(i)$ ($n = 5, 10, 20, 30, 50$ and 100)

The rth-order moment of random variable X with the probability distribution (2) is given by the following recurrence relation (Tsuchiya and Nakamura, 2009):

$$E_n(X^r) = \begin{cases} 1 & (n = 1) \\ \displaystyle\sum_{k=0}^{r} \left\{ \frac{n+1}{n} \binom{r}{k} - \frac{1}{n} \binom{r+1}{k} \right\} E_{n-1}(X^k) & (n \geq 2). \end{cases} \tag{3}$$

The expectation and the variance of random variable X with the probability distribution (2) are respectively given by

$$E_n(X) = \frac{n+1}{2} \quad (n \geq 1),$$

$$V_n(X) = \begin{cases} 0 & (n = 1) \\ \dfrac{n+1}{12} & (n \geq 2). \end{cases}$$

5. Eulerian distribution with a missing number

By extending the above sorting process, we derive the probability distribution of the number of bunches in the case that there exists a missing number in a set of cards. We investigate here the probability distribution with just one missing number. The probability distribution is derived under the assumption that any one of the cards is missing number with equal probability.

For example, when $n = 3$, if there are not any missing numbers, then we see that the cases that the cards appear are in the following six sequences:

$$(1,2,3), \quad (1,3,2), \quad (2,1,3), \quad (2,3,1), \quad (3,1,2), \quad (3,2,1). \tag{4}$$

The situation that just one of the cards is missing number is equivalent to the case that the third number is missed in each arrangement of (4). Then, we have the following: $(1,2)$, $(1,3)$, $(2,1)$, $(2,3)$, $(3,1)$, $(3,2)$. Further, the number of bunches decreases by one for two cases $(1,2)$, $(2,1)$. They occur when the third number is 3 in (4) and the number is missed. Then it should be noted that the number of bunches in the remaining arrangements is equal to the case of $n = 2$. For the other cases, the number of bunches is invariant under this operation. Therefore, the total number of the case that the number of bunches $X = i$, denoted by $M_3^{(1)}(i)$ can be obtained by adding $M_3(i)$ to $M_2(i)$ and subtracting $M_2(i-1)$ from that. Similarly, when $n = 4$, if the fourth number is 4 and the number is only missed, then the number of bunches decreases by one. Hence, we obtain

$M_4^{(1)}(i) = M_4(i) + M_3(i) - M_3(i-1)$.

Let $M_n^{(1)}(i)$ be the total number of the case that the number of bunches $X = i$ in n cards with a missing number. From the above argument, the following recurrence relation holds:

$$\begin{cases} M_n^{(1)}(1) = 2 & (n \geq 2) \\ M_n^{(1)}(n) = 0 & (n \geq 1) \\ M_n^{(1)}(i) = M_n(i) + M_{n-1}(i) - M_{n-1}(i-1) & (n \geq 3,\ i = 2, \cdots, n-1), \end{cases}$$

where $M_n(i)$ is the total number of the case for bunches without missing number and satisfies the recurrence relation (1). The values of $M_n^{(1)}(i)$ $(1 \leq i < n \leq 10)$ are given by Table 2.

Table 2. $M_n^{(1)}(i)$ $(n = 2, 3, \cdots, 10)$

$n \backslash i$	1	2	3	4	5	6	7	8	9
2	2								
3	2	4							
4	2	14	8						
5	2	36	66	16					
6	2	82	342	262	32				
7	2	176	1436	2416	946	64			
8	2	366	5364	16844	14394	3222	128		
9	2	748	18654	99560	156190	76908	10562	256	
10	2	1514	61946	528818	1378310	1242398	381566	33734	512

Then, the probability that $X = i$ can be obtained as $P_n^{(1)}(i) = M_n^{(1)}(i)/n!$ and is given by the following recurrence formula:

$$\begin{cases} P_n^{(1)}(1) = \dfrac{2}{n!} & (n \geq 2) \\ P_n^{(1)}(n) = 0 & (n \geq 1) \\ P_n^{(1)}(i) = P_n(i) + \dfrac{1}{n}P_{n-1}(i) - \dfrac{1}{n}P_{n-1}(i-1) & (n \geq 3,\ i = 2, \cdots, n-1), \end{cases} \tag{5}$$

where $P_n(i)$ satisfies the recurrence relation (2). Figure 2 shows the probability distribution $P_n^{(1)}(i)$ for $n = 5, 10, 20, 30, 50$ and 100. It can be seen that the distribution is not symmetric with respect to the mean. But we find that the distribution approaches to a normal distribution as n becomes large.

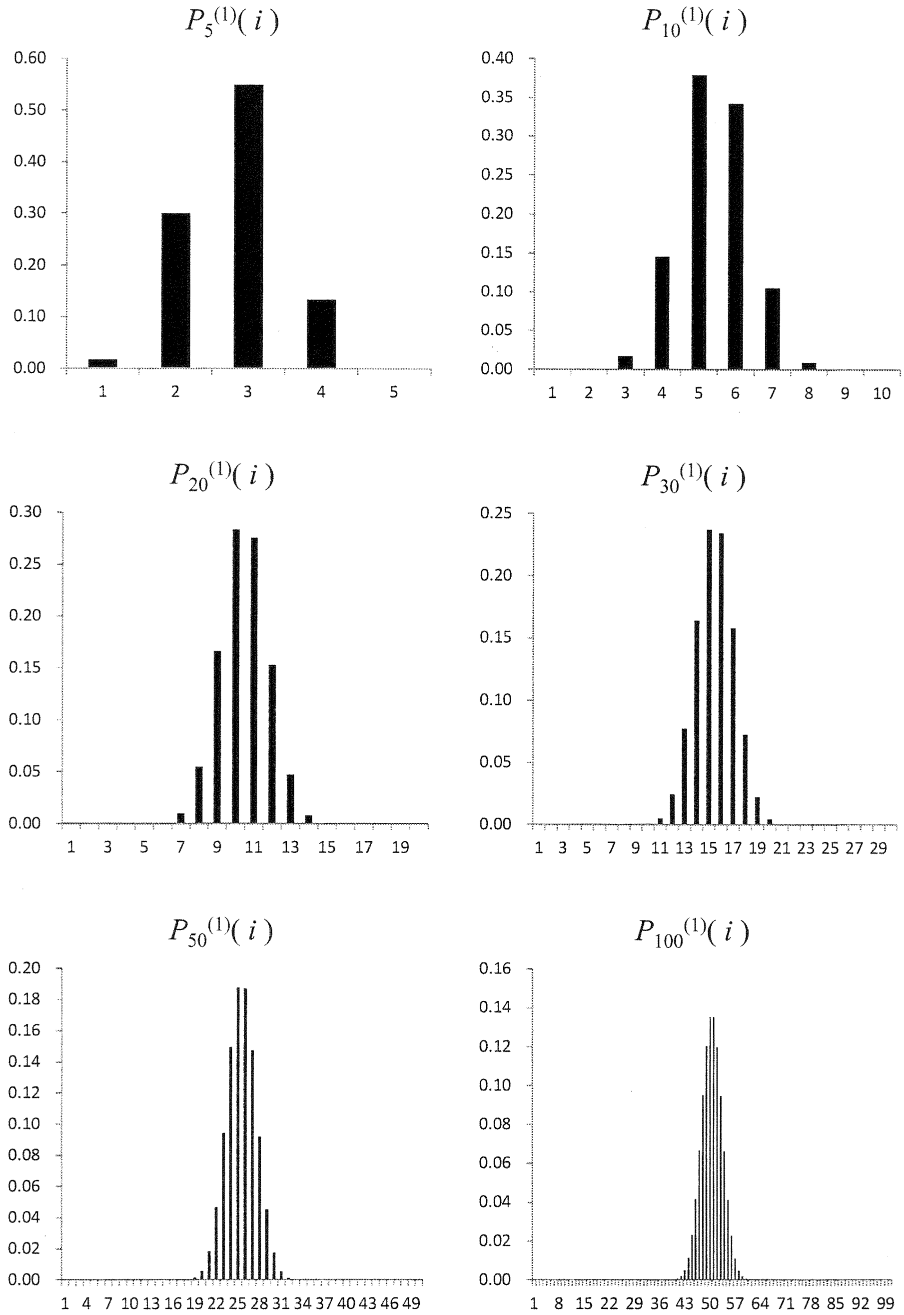

Figure 2. $P_n^{(1)}(i)$ ($n = 5, 10, 20, 30, 50$ and 100)

THEOREM 1. *The rth-order moment of random variable X with the probability distribution (5) is given by the following recurrence relation:*

$$E_n^{(1)}(X^r) = \begin{cases} 0 & (n=1) \\ E_n(X^r) - \dfrac{1}{n}\displaystyle\sum_{k=0}^{r-1}\binom{r}{k} E_{n-1}(X^k) & (n \geq 2), \end{cases} \tag{6}$$

where $E_n(X^r)$ satisfies the recurrence relation (3).

Proof. By definition of the rth-order moment, $E_1^{(1)}(X^r) = 1^r P_1^{(1)}(1) = 0$ always holds for $n = 1$. From (5), for $n \geq 3$, we have the following equation:

$$\sum_{i=2}^{n-1} i^r P_n^{(1)}(i) = \sum_{i=2}^{n-1} i^r P_n(i) + \frac{1}{n}\sum_{i=2}^{n-1} i^r P_{n-1}(i) - \frac{1}{n}\sum_{i=2}^{n-1} i^r P_{n-1}(i-1). \tag{7}$$

Since $1^r P_n^{(1)}(1) = 2/n!$, $n^r P_n^{(1)}(n) = 0$, $1^r P_n(1) = 1/n!$ and $1^r P_{n-1}(1) = 1/(n-1)!$, the equation (7) is represented by

$$\sum_{i=1}^{n} i^r P_n^{(1)}(i) = \sum_{i=1}^{n-1} i^r P_n(i) + \frac{1}{n}\sum_{i=1}^{n-1} i^r P_{n-1}(i) - \frac{1}{n}\sum_{i=2}^{n-1} i^r P_{n-1}(i-1). \tag{8}$$

The third term on the right-hand side in (8) can be written as

$$\frac{1}{n}\sum_{i=2}^{n-1} i^r P_{n-1}(i-1) = \frac{1}{n}\sum_{i=1}^{n-2} (i+1)^r P_{n-1}(i).$$

Further, we have $n^r P_n(n) = n^r/n!$ and $n^r P_{n-1}(n-1) = n^r/(n-1)!$. Therefore, (8) is expressed as

$$\sum_{i=1}^{n} i^r P_n^{(1)}(i) = \sum_{i=1}^{n} i^r P_n(i) + \frac{1}{n}\sum_{i=1}^{n-1} i^r P_{n-1}(i) - \frac{1}{n}\sum_{i=1}^{n-1} (i+1)^r P_{n-1}(i).$$

Hence, the recurrence relation with respect to rth-order moment $E_n^{(1)}(X^r)$ is given by

$$\begin{aligned} E_n^{(1)}(X^r) &= E_n(X^r) + \frac{1}{n}E_{n-1}(X^r) - \frac{1}{n}E_{n-1}[(X+1)^r] \\ &= E_n(X^r) + \frac{1}{n}E_{n-1}(X^r) - \frac{1}{n}\sum_{k=0}^{r}\binom{r}{k} E_{n-1}(X^k) \end{aligned}$$

$$= E_n(X^r) - \frac{1}{n}\sum_{k=0}^{r-1}\binom{r}{k} E_{n-1}(X^k). \tag{9}$$

From (3), we have

$$\begin{aligned} E_2(X^r) - \frac{1}{2}\sum_{k=0}^{r-1}\binom{r}{k} E_1(X^k) &= \frac{3}{2}\sum_{k=0}^{r}\binom{r}{k} - \frac{1}{2}\sum_{k=0}^{r}\binom{r+1}{k} - \frac{1}{2}\sum_{k=0}^{r-1}\binom{r}{k} \\ &= \frac{3}{2}\cdot 2^r - \frac{1}{2}(2^{r+1}-1) - \frac{1}{2}(2^r - 1) \\ &= 1. \end{aligned}$$

On the other hand, since it follows from (5) that

$$E_2^{(1)}(X^r) = 1^r P_2^{(1)}(1) + 2^r P_2^{(1)}(2) = 1,$$

we see that the equation (9) also holds for $n = 2$. □

COROLLARY 2. *The expectation and the variance of random variable X with the probability distribution (5) are respectively given by*

$$E_n^{(1)}(X) = \frac{n+1}{2} - \frac{1}{n} \quad (n \geq 1), \tag{10}$$

$$V_n^{(1)}(X) = \begin{cases} 0 & (n=1) \\ \dfrac{n+1}{12} - \dfrac{1}{n^2} & (n \geq 2). \end{cases} \tag{11}$$

Proof. By putting $r = 1$ in (6), the expectation of X is given by

$$\begin{aligned} E_n^{(1)}(X) &= E_n(X) - \frac{1}{n}E_{n-1}(1) \\ &= \frac{n+1}{2} - \frac{1}{n} \end{aligned} \tag{12}$$

for $n \geq 2$. Since $E_1^{(1)}(X) = 0$, the equation (12) also holds for $n = 1$. Further, by putting $r = 2$ in (6), the expectation of X^2 can be expressed as

$$\begin{aligned} E_n^{(1)}(X^2) &= E_n(X^2) - \frac{1}{n}\{E_{n-1}(1) + 2E_{n-1}(X)\} \\ &= E_n(X^2) - \frac{1}{n}(1+n) \end{aligned}$$

for $n \geq 2$. Therefore, the variance of X is given by

$$\begin{aligned} V_n^{(1)}(X) &= V_n(X) - \frac{1}{n^2} \\ &= \frac{n+1}{12} - \frac{1}{n^2} \end{aligned}$$

for $n \geq 2$ and $V_n^{(1)}(X) = 0$ for $n = 1$. □

6. Numerical comparisons

Table 3 shows the results of comparing the sample mean and the unbiased variance based on random samples of sizes $n = 5, 10, 20, 30, 50$ and 100 drawn from the distribution (5) with the expectation and the variance by using (10) and (11). The simulation results were obtained by averaging over 10,000 repeated random samples. It can be seen from the table that the results in simulation is almost the same as theoretical values.

Table 3. Comparison of means and variances in simulation study and theoretical values for various values of n.

	Simulation results		Theoretical values	
n	Sample mean	Unbiased variance	Expectation	Variance
5	2.80	0.450	2.80	0.460
10	5.40	0.911	5.40	0.907
20	10.45	1.762	10.45	1.748
30	15.45	2.628	15.47	2.582
50	25.48	4.249	25.48	4.250
100	50.53	8.486	50.49	8.417

7. Conclusion

In this paper, we derived the recurrence relation for Eulerian distribution with a missing number and the moment of the probability distribution. We see that the distribution is non-symmetric and the expectation and variance, respectively, decrease by $1/n$ and $1/n^2$, comparing with the distribution which does not have any missing number. The remaining problem is to reveal the relation of the derived distribution and the other probability distribution, and the approximation to the distribution and to derive the probability distribution with some missing numbers.

Acknowledgments. I would like to thank Professor Nagatomo Nakamura and the referees for valuable suggestions and comments. This research was supported by the ISM Cooperative Research Program (2014-ISM・CRP-2067).

References

[1] Euler, L. (1755). *Institutiones Calculi Differentialis*. St. Petersburg. *Foundations of Differential Calculus*, John D. Blanton (translator), Springer, New York, 2000.

[2] Graham, R. L., Knuth, D. E. and Patashnik, O. (1994). Eulerian Numbers. *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed., Section 6.2, 267-272, Addison-Wesley, Boston.

[3] Mann, H. B. (1945). On a test for randomness based on signs of differences. *Annals of Mathematical Statistics*, **16**, 193-199.

[4] Nakamura, N. (2015). Pseudo-Normal Random Number Generation via the Eulerian Numbers. *Josai Mathematical Monographs*, **8**, 85-95.

[5] Tsuchiya, T. and Nakamura, N. (2007). A probability distribution function with modified backet sort (in Japanese). *Sapporo Gakuin University review of economics and business*, **4**(2), 31-47.

[6] Tsuchiya, T. and Nakamura, N. (2009). Eulerian numbers in modified bucket sorting and its related distribution theories (in Japanese). *Proceedings of the Institute of Statistical Mathematics*, **57**(1), 159-178.

Takahiro Tsuchiya

Department of Mathematics, Faculty of Science, Josai University

Keyakidai 1-1, Sakado, Saitama, 350-0295, Japan

takahiro@josai.ac.jp